In special issue of *Neural Networks* (selected papers from 2003 International Joint Conference on Neural Networks)

Multimedia Authenticity Protection With ICA Watermarking and Digital Bacteria Vaccination

Harold Szu[1], Steven Noel[2], Seong-Bin Yim[3], Jeff Willey[4], and Joe Landa[5]
[1]Office of Naval Research, 800 N. Quincy St., Arlington VA 22217-5660
[2]Center for Secure Information Systems, George Mason University, VA 22030-4444
[3]Digital Media RF Lab, ECE Dept., The George Washington University, Washington DC 20052
[4]Code 5344, Naval Research Laboratory, Washington, DC 20375-5336
[5]BriarTek, Inc., 112 E. Del Ray Ave., Suite A, Alexandria, VA 22301

***Abstract*—We propose the application of independent component analysis (ICA), via unsupervised neural networks, to authenticity protection for multimedia products. We give an overview of the current state of multimedia authenticity protection, including the requirements of various multimedia applications, current approaches to the problem, and the robustness of the approaches. For watermark security, a covert independent-component watermarking signal can serve as a "vaccination" against a dormant digital "bacteria" protecting the multimedia data. An unauthorized removal of the watermark triggers the bacteria payload, which then degrades the quality of the unauthorized data. We argue that such digital bacteria meet all the established requirements for beneficial virus-like programs, and their payload would merely affect pirated media. We show how these new approaches contribute to a flexible, robust, and secure system for protecting the authenticity of multimedia products.**

***Keywords*—Multimedia watermark, copyright protection, Internet commerce, independent component analysis, unsupervised neural networks.**

I. Introduction

The ubiquitous piracy of copyrighted digital music files over the Internet threatens the livelihood of content-providing businesses as well as the creative artists who hold the copyright. Other than the legal and practical issues related to copyright enforcement, several technology factors influence the piracy. The major technology component is the presence of easy to use Internet based peer-to-peer file-sharing technology. Another factor is the compact format and high fidelity of Moving Picture Expert Group-I (MPEG-I) Layer-3 (MP3) digitally encoded music files. The result is that peer-to-peer MP3 encoded music files can be obtained quickly over high-speed network connections.

The application of Independent Component Analyses (ICA) to digital watermarking, for the purpose of multimedia authenticity protection, was first introduced in (Noel and Szu, 2000). This has foundations in previous work in ICA for intelligent sensory processing, i.e., (Szu, 1999a), (Szu, 1999b), and (Quan, Szu, and Markovitz, 2000).

Many other technical strategies have been both proposed and pursued to reduce the piracy problem. One approach is to embed a digital code or watermark in the music itself. The watermark can be chosen to be imperceptible to the listener or deliberately chosen to corrupt the musical audio signal.

The ideal watermarking technique for digital music files has several desirable attributes including (1) compatibility with the popular MP3 format, (2) resistance to watermark removal by the consumer, and (3) complete recovery of the music from the watermark with authorized/restricted computer codes. We explore the efficacy of a novel approach (linear mixing and blind demixing with ICA) for watermarking digital music files to fulfill these goals.

Earlier research on watermarking via ICA concentrated on either detection, (Szu, 1999a), (Szu, 1999b), (Hartung and Kutter, 1999), (Swanson, Kobayashi, and Tewfick, 1998), and (Jessop, 1999), or on application to imagery, (Noel and Szu, 2000), (Quan, Szu, and Markovitz, 2000), (Kopriva and Szu, 2003), (Barnett, 1999), and (Seok and Hong, 2001). Instead, this paper focuses on some of the subtle and practical issues of ICA for watermarking digital music files, especially robustness with respect to noisy MP3 compression.

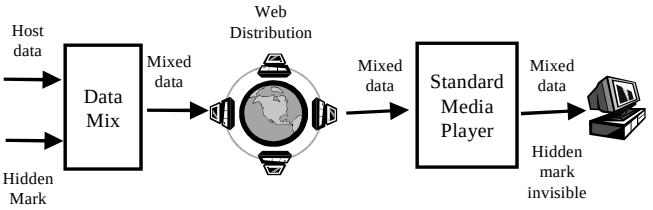

Fig. 1. Standard watermarking model.

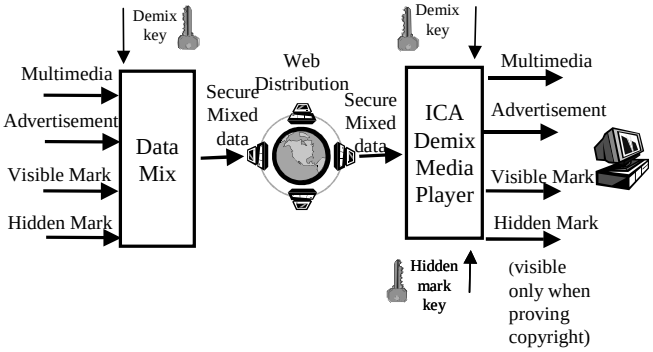

Fig. 2. ICA demix enabled model.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUL 2003		**00-00-2003 to 00-00-2003**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Multimedia Authenticity Protection With ICA Watermarking and Digital Bacteria Vaccination	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Office of Naval Research,800 N. Quincy St,Arlington,VA,22217-5660**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
2003 International Joint Conference on Neural Networks, 20-24 July, Portland, OR

14. ABSTRACT
We propose the application of independent component analysis (ICA), via unsupervised neural networks, to authenticity protection for multimedia products. We give an overview of the current state of multimedia authenticity protection including the requirements of various multimedia applications current approaches to the problem, and the robustness of the approaches. For watermark security, a covert independentcomponent watermarking signal can serve as a ?vaccination? against a dormant digital ?bacteria? protecting the multimedia data. An unauthorized removal of the watermark triggers the bacteria payload, which then degrades the quality of the unauthorized data. We argue that such digital bacteria meet all the established requirements for beneficial virus-like programs, and their payload would merely affect pirated media. We show how these new approaches contribute to a flexible, robust, and secure system for protecting the authenticity of multimedia products.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

Earlier work (Noel and Szu, 2000) described two models for watermarking of multimedia products: (1) a standard model and (2) a model for neural net ICA demix enabled media player (Fig.1 and Fig. 2, respectively). The model in Fig. 1 uses a standard media player; therefore, the hidden watermark is invisible. On the other hand, the model in Fig. 2 uses an ICA demixed media player to recover the hidden watermark when proving copyright. Previous work (Noel and Szu, 2000) points out a critical issue: ICA needs to be able to demix signals that have been subjected to lossy compression. It also demonstrates that ICA neural networks can blindly demix media signals with relatively little distortion after the application of lossy wavelet compression.

In this paper, the idea is to nonlinearly mix a watermark audio signal with the music signal, MP3 encode and decode the resulting signal, then attempt to demix the watermark signal from the music. The novelty of this method introduces difficulty in the removal of the watermark by the unsophisticated consumer. The value of restricted computer codes to restore the music from a watermarked signal will depend in part on the quality of music after blind demixing with ICA (Yu, Satter, and Ma, 2002), (Yu and Satter, 2002), (Gonzalez-Serrano, Molina-Bulla, and Murillo-Fuentes, 2001), (Bell and Sejnowski, 1995), (Bell and Sejnowski, 1996), (Amari, Chichocki, and Yang, 1996). We measure the watermark contamination of the music after blind demixing with ICA with MP3 encoding.

In ICA, sensed signals are modeled as statistically independent components, which are then linearly mixed. Here independence is over all orders of statistics, not just 2^{nd}-order correlations. Neural networks with unsupervised learning rules are able to estimate the signal mixing to high accuracy, so that the original signals (independent components) can be recovered through inverse mixing.

The ability to blindly demix signals enables a novel form of security against those who may attack multimedia watermarks. One of the independent component signals in the mixed data stream could be a copy of an overt (visible or audible) authenticity mark on the host signal. A multimedia player enabled with an ICA neural network could then blindly demix the overt mark copy and determine whether the mark is still present in the host data.

If the overt authenticity mark has been removed, then a program triggers that degrades the quality of the unauthorized data. We call this program an electronic "bacterium," as opposed to a "virus," in the sense that it does not replicate indiscriminately. Moreover, there exists a "vaccine" against this electronic bacterium, i.e., the presence of the authenticity mark. We also investigate two key requirements of the approach: (1) the robustness of ICA digital watermarking with respect to lossy compression and (2) the dynamic range of the digital watermark as clutter versus original music, for a signal coded under the MPEG-3.

In the next section, we review ICA theory and introduce the concept of multimedia watermarks via ICA. Section III then applies ICA watermarks within a full framework for protecting multimedia authenticity. Sections IV through VII then perform experiments testing the robustness of the approach for MPEG-3 audio.

II. Blind Demixing of Multimedia Data with ICA Neural Networks

Existing watermarking systems are essentially non-blind, in the sense that they require original versions of both the host data and watermark keystream in order to extract the watermark. This may be reasonable for applications that merely verify the existence of the watermark for proving ownership. However, non-blind schemes are inadequate for our application, in which embedded data is extracted on the consumer side, not the producer side.

A straightforward way to combine host and watermark signals is to combine them linearly. The signals could then later be demixed given the linear mixing coefficients. However, this provides little in the way of flexibility or security. It would be much better if the signals could be demixed without knowledge of the original mixing coefficients, or even of the original signals themselves. Thus, we are faced with the problem of blind demixing of source signals.

In ICA, the source signals are modeled as statistically independent signal components, which have been subsequently mixed linearly. Independence is defined such that the joint probability densities of the signal components can be factorized as the product of the marginal densities. Thus, independence is over all orders of statistics, not just 2^{nd}-order correlations.

Neural networks with unsupervised learning are able to estimate the signal mixing to high accuracy, so that the original signals (independent components) can be recovered through inverse mixing. The learning rules are based on maximizing the degree to which the independent components are non-gaussian. Possible measures of nongaussianity are the absolute value of kurtosis (4^{th}-order cumulant), or negentropy (slightly modified version of differential entropy).

ICA assumes that there are multiple sensors, each sensing a different mix of the independent components. Thus is extends conventional single-sensor processing of signals to multiple sensors, analogous to the multiple sensors (eyes and ears) in humans. The unsupervised ICA neural networks are able to simultaneously compare sensor outputs, extracting noise so that only coherent signals remain. In the unsupervised learning rule, there is no specific desired output other than white gaussian noise. This is consistent with the idea that what is not noise must be signals.

More formally, assume n mutually independent signal sources at each time instant, represented as the vector $\boldsymbol{s}=[s_1(t), s_2(t),\ldots, s_n(t)]^T$, which are then linearly mixed with an unknown mixing matrix $\boldsymbol{A}$ onto a vector of received signals $\boldsymbol{x}=[x_1(t), x_2(t),\ldots,x_m(t)]^T$, i.e.,

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s}, \tag{1}$$

where $\boldsymbol{A}$ is an $m \times n$ scalar matrix of full rank. Now let a neural processor with weight matrix $\boldsymbol{W}$ and non-linear contrast function g form the output $\boldsymbol{y}$:

$$\boldsymbol{u} = \boldsymbol{W}\boldsymbol{x}, \tag{2}$$

$$\boldsymbol{y} = g(\boldsymbol{u}), \tag{3}$$

where $\boldsymbol{W}$ is an $n \times m$ scalar weight matrix. Further, let the following additional assumptions hold:

1. The function g_i, is the cumulative distribution function of the source probability density function, $p(s_i(t))$.
2. The number of received signals, m, is at least equal to the number of source signals, n.
3. At most one source is normally distributed.
4. The receiver noise is negligible relative to the source signal power.

When speed of media processing is important, we recommend pixel-parallel Lagrange Constraint Neural Networks (LCNN), which solves blind source separation pixel by pixel in parallel. It has been shown (Szu, 1999a), (Szu, 1999b) that minimization of Helmholtz free energy $\boldsymbol{H} = \boldsymbol{E} - \boldsymbol{T}_o \boldsymbol{S}$ will achieve blind signal source separation, where $\boldsymbol{E}$ is the Lagrange constraint first-order estimation error

$$\boldsymbol{E} = \lambda([\boldsymbol{A}]\boldsymbol{s} - \boldsymbol{x}) = \mu([\boldsymbol{W}]\boldsymbol{x} - \boldsymbol{s}) .$$

Here T_o is the constant temperature of thermal reservoir, and S is the Shannon Boltzmann entropy. Two LCNN theorems are reviewed as follows, the first one without data, the other with data.

Theorem 1. An equal partition law at Maximum Entropy: Without being biased by pixel data, the equilibrium distribution at the maximum of entropy is an ***equal partition law*** as follows:

$$\boldsymbol{S} = -\boldsymbol{K}_B \sum_{j=1,M} \boldsymbol{s}_j \log \boldsymbol{s}_j + \boldsymbol{K}_B(\lambda_o + 1)(\sum_{j=1,M} \boldsymbol{s}_j - 1) .$$

Next, $\frac{\partial \boldsymbol{S}}{\partial \boldsymbol{s}_j} = -\boldsymbol{K}_B(\log \boldsymbol{s}_j + 1) + \boldsymbol{K}_B(\lambda_o + 1) = 0$ yields

$\boldsymbol{s}_j = \exp(\lambda_o)$, and imposing the unit sum $\sum_{j=1,M} \boldsymbol{s}_j = 1$ we find $\boldsymbol{s}_j = \frac{1}{\boldsymbol{M}}$ for the equal partition distribution. **Q.E.D.**

By computation of the Amari natural gradient with respect to the weight matrix $\boldsymbol{W}$, and the distance metric $[A]^T[A]$, we have

$$\mathrm{dx} = (\mathrm{x}, \mathrm{x}) = (\mathrm{s}, [\mathrm{A}]^T[\mathrm{A}]\mathrm{s}) = \mathrm{d}$$

$$\frac{\partial \boldsymbol{A}}{\partial \boldsymbol{t}} = \frac{\partial \boldsymbol{H}(\boldsymbol{x})}{\partial \boldsymbol{A}} \boldsymbol{A}^T \boldsymbol{A} , \tag{4}$$

Theorem 2. ANN Sigmoid Distribution

Setting the derivative to zero for the minimum solution $\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{s}_j} = \sum_k \lambda_k \boldsymbol{A}_{kj} + \log \boldsymbol{s}_j + 1 - (\lambda_o + 1) = 0$ reproduces, by the unit sum constraints $\sum_j \boldsymbol{s}_j = 1$, the ANN sigmoid threshold (without assuming it for ICA post processing):

$$\boldsymbol{s}_j = \exp(\frac{\lambda_j}{\boldsymbol{K}_B \boldsymbol{T}_o} + \lambda_o)$$

$$= \frac{1}{1 + \sum_{k \neq j} \exp(\frac{\lambda_k}{\boldsymbol{K}_B \boldsymbol{T}_o} - \frac{\lambda_j}{\boldsymbol{K}_B \boldsymbol{T}_o})} = \sigma(\frac{\lambda_j}{\boldsymbol{K}_B \boldsymbol{T}_o})$$

Q.E.D.

The Lagrange constraint vector could be obtained with a standard textbook in physics. Thus, we achieved single pixel blind source separation, i.e., the output would not have the pixel ensemble average led to a permutation and scaled version of the original mutually independent signal sources $\boldsymbol{s}$.

On the other hand, a batch-mode operation uses the feed forward weight update rule (Bell and Sejnowski, 1995) to have the general form

$$\frac{\partial \boldsymbol{S}(\boldsymbol{y})}{\partial \boldsymbol{W}} = (\boldsymbol{W}^T)^{-1} + \left(\frac{\frac{\partial \boldsymbol{p}(\boldsymbol{u})}{\partial \boldsymbol{u}}}{\boldsymbol{p}(\boldsymbol{u})} \right) \boldsymbol{x}^T , \tag{5}$$

Here $p(u)$ is a source probability density function. As (Amari, Chichocki, and Yang, 1996) shows by post multiplication of (5) by $\boldsymbol{W}^T\boldsymbol{W}$, one has

$$\frac{\partial \boldsymbol{S}(\boldsymbol{y})}{\partial \boldsymbol{W}} = \left[\boldsymbol{I} + \left(\frac{\frac{\partial \boldsymbol{p}(\boldsymbol{u})}{\partial \boldsymbol{u}}}{\boldsymbol{p}(\boldsymbol{u})} \right) \boldsymbol{u}^T \right] \boldsymbol{W} . \tag{6}$$

In practice, we employ fixed-point Fast ICA Matlab code (Oja and Karhunen, 1995) to compute the gradient. Here the hyperbolic tangent function is chosen for the contrast related function, $(d\,p(u)/\,d\,u)/\,p(u)$, in (6).

Fig. 3 is an example of the mixing of multimedia data, with blind demixing by an ICA unsupervised neural network. The three independent data components are the host data, an advertisement, and an authenticity mark. The data are mixed to insure the various components remain together for Internet distribution. The neural network with unsupervised learning estimates with high accuracy the mixing matrix **A**, allowing blind demixing of the independent components.

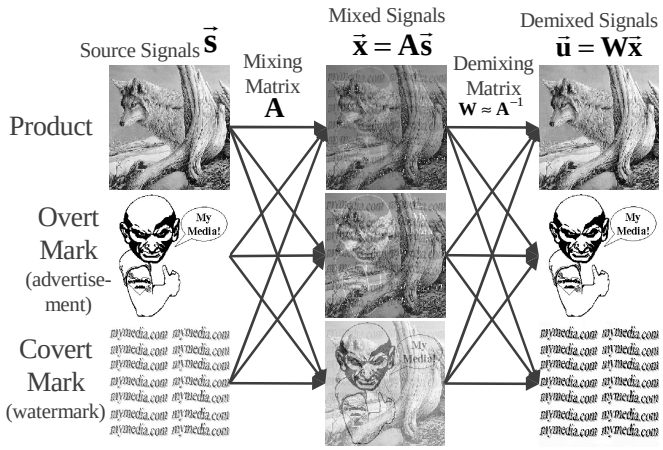

Fig. 3. Example mixing of multimedia data, with blind demixing via ICA unsupervised neural network.

III. Independent Components for Multimedia Watermark Vaccination

It is important that ICA demixing be robust with respect to various forms of signal processing. Being linear, ICA demixing is invariant with respect to linear filtering. It is also invariant with respect to various changes in sampling, which cause synchronization problems with most other watermarking techniques. Such sampling changes include cropping, line dropping, or changing the sample rate. In the ICA model, each sample point is a mix of two independent components, and the mix is same for all samples.

It is also critical that ICA neural networks be able to demix signals that have been subjected to lossy compression. In general, compression can occur either before or after mixing. Compression before mixing yields nearly lossless demixing, and avoids costly decompression of previously compressed data. As we see in Fig. 4 for audio data, ICA neural networks can blindly demix signals with minimal distortion even after they have been subjected to the nonlinear transformation of lossy wavelet compression.

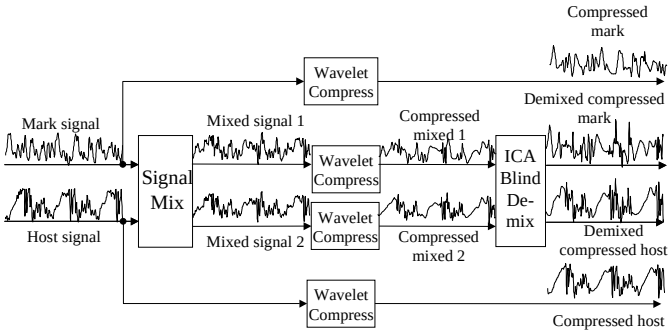

Fig.4. Robustness of ICA neural net blind demixing with respect to nonlinear wavelet audio data compression.

In Fig. 5, after the host and watermark images are mixed, the mixtures are subjected to DWT compression. ICA unsupervised neural nets are still able to blindly demix host and mark images, despite the nonlinear transformation of mixed images via compression. In fact, there is little visual difference compared to the compressed versions with no mixing involved.

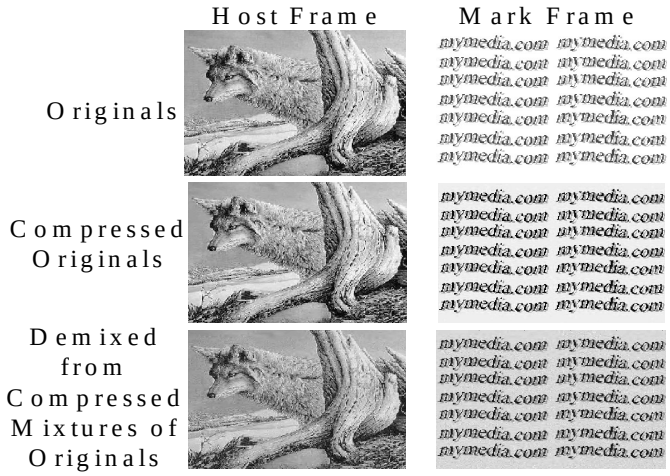

Fig.5. Robustness of ICA neural net blind demixing with respect to nonlinear wavelet image data compression.

Once the data is demixed, it is open to attack by the consumer. For example, the consumer could attempt to remove a visible overt company logo from the demixed multimedia product, to claim ownership. We do not have the luxury of a DVD-like approach, where there is industry-wide cooperation and all multimedia players are trusted. Our approach needs to be completely autonomous.

Our approach to making product authenticity secure against someone who has the demix key is essentially a watermark "vaccination" against digital "bacteria." The purpose of the bacteria is to degrade the quality of the multimedia product if it has been tampered with, e.g. its company logo is removed.

The vaccination is the actual presence of the company logo, and it has already been administered in the sense that the logo is initially part of the product. Hidden data localized to the company logo serves as the flag for the logo's presence. The level of infection (i.e., the action to take upon logo deletion) can vary from doing nothing, to degrading sound/video quality, to rendering the multimedia product unusable, depending on the choice of product owner. The infection is localized to the unauthorized multimedia file, and spreads to no other part of computer, since unauthorized content is directly recognizable via ICA demixing.

Fred Cohen pioneered research in computer viruses [1] (Cohen, 1987). His early results showed that (1) viruses could spread unhindered, even in secured networks, (2) that they could cause essentially unlimited damage with little effort by the virus writer, (3) that virus detection was undecid-

[1] In the early 1980s, as a Ph.D. student at the University of Southern California, Fred Cohen got the idea of self-replicating software that spreads by attaching itself to existing programs. He shared this idea with his thesis advisor Len Adleman, who pointed out the similarity to a biological virus, leading to the term "computer virus."

able, and (4) that many of the defenses that could be devised relatively quickly were ineffective against a serious virus writer. Despite subsequent advances in virus defense, it appears that it will always be possible to write effective viruses.

For various reasons, the general public usually considers computer viruses to be malicious code with no useful purpose. However, maliciousness is neither a necessary nor a sufficient property for computer viruses. Indeed, Cohen has described the potential advantages of beneficial viruses.

It is true that there have been relatively few successful examples of beneficial viruses, despite several attempts. Bontchev argues convincingly (Bontchev, 1994) that such failed attempts violate one or more important properties that beneficial viruses should have.

In particular, the technical properties that beneficial viruses should have are (1) control, (2) recognition, (3), no resource wasting, (4) bug containment, (5) compatibility, and (6) effectiveness. Such viruses should also meet ethical/legal requirements with respect to unauthorized data modification and copyrights, as well as psychological barriers involving lack of trust and negative connotations with computer viruses.

Bontchev also describes a general model for beneficial viruses that satisfies these requirements (Bontchev, 1994). This model requires the active consent of individual users, via cryptographically strong system authentication. It also requires that the virus be self-contained and propagate as a whole, and not depend on attaching itself to a host executable file. The beneficial virus must also consume negligible resources, at least compared to the benefits it provides.

Our proposed digital bacteria fit within this general model for beneficial viruses. They fulfill the control requirement because they only spread to machines that have agreed to the network policy with respect to unauthorized media. They fulfill the recognition property because they can be easily ignored by anti-virus scanners, since they can authenticate themselves as a known beneficial virus.

Our digital bacteria meet the resource-wasting criterion, since only a single instance of the virus is present on an infected machine, and assuming the media degradation process is efficient. Through public-key cryptography for authentication, digital bacteria updates can be strictly controlled, addressing the bug-containment problem.

Because the digital bacteria are self-contained and do not modify other programs, they introduce no greater risk on incompatibility than non-viral programs. Our digital bacteria meet the effectiveness requirement, since a virus-like propagation is a highly effective way to ensure multimedia authenticity protection across a participating network.

Since our proposed digital bacteria modify only unauthorized media files, there are no ethical/legal problems. The psychological barrier is largely overcome by the fact that most people would not consider such programs to be actual viruses. Some users may resent the effects of the digital bacteria if they thwart their efforts at illegally copying music or videos. But network administrators would appreciate that illegal activities are being curtailed on their network, and there would be definite monetary incentives by the media content producers and artists.

One way to implement watermark vaccination is to have the vaccination-checking executable code downloaded along with the mixed data and multimedia player. It is then executed along with the player, as shown in Fig. 8. Multimedia data, advertisement, and mark data are mixed and secured, then embedded within the object. The neural network ICA demix media player is defined as an object method, with read-only access to mixed data. The demix key allows data to be demixed, and the mark key allows the hidden mark to be accessed for copyright proof.

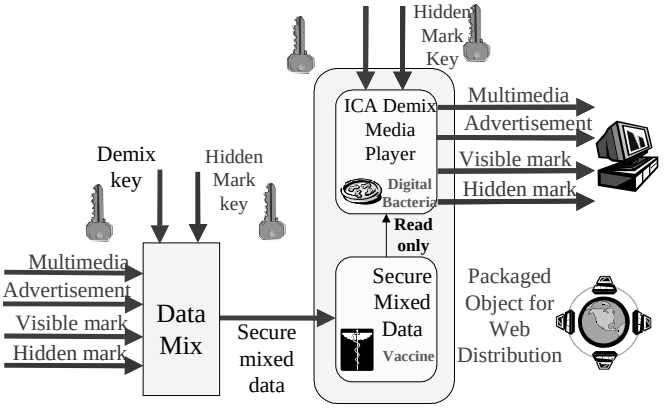

Fig. 8. Object-based model for multimedia product protection by watermark vaccination.

Another implementation is DVD-like, that is, via trusted players. Either the multimedia producers or a third party provides the player/protection software cost-free. This approach requires universal participation for complete protection; otherwise, there is protection only among trusted players. A third approach is to implement a DVD-like scheme in hardware, with an interface to the operating system.

IV. Experimental Background

A. Blind Demixing with ICA

The idea is to linearly mix a watermark acoustic signal with the music signal, MP3 encode and decode the resulting signal, then attempt to demix the watermark signal from the music. The novelty of this method introduces difficulty in the removal of the watermark by the unsophisticated consumer. The value of restricted computer codes to restore the music from a watermarked signal will depend in part on the quality of music after blind demixing with ICA. We measure the watermark contamination of the music after blind demixing with ICA with MP3 encoding. The receiver noise is negligible relative to the source signal power.

B. MP3 Encoding/Decoding

The compact MP3 encoding format is a lossy compression method. It utilizes a psycho-acoustic model of human hearing to discard information (Pan, 1995), (Brandenburg and Popp, 2000). As shown in Fig. 9, human hearing is subject to

auditory masking, where small tones in the local frequency neighborhood of a much stronger tone are imperceptible.

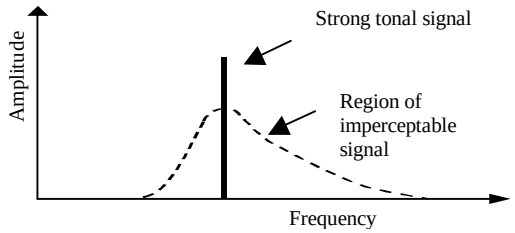

Fig. 9. Illustration of audio noise masking. Weak signals in the local frequency neighborhood of a strong tone are imperceptible.

Further, the human auditory system's frequency resolution is a function of absolute frequency; perceptible filter widths are narrow at the low end and significantly wider at the high end. Thus, weak tones adjacent to stronger tones may be eliminated by MP3 encoding/decoding.

Several issues arise when considering ICA for blind demixing of MP3 watermarked music. First, introduction of a watermark signal must be considered in the context of the music, so that it will not be eliminated by the lossy MP3 encoding/decoding. A second issue is consideration of the statistics of the music and watermark signals after MP3 encoding/decoding. The concern is that the statistics of both the watermark and music signals may be perturbed when selected frequency components are nulled. The altered statistics may adversely affect the ICA blind demixing.

Our strategy is to measure the watermark contamination in the music signal after blind demixing with ICA. First we measure the watermark residue in the music channel without MP3 encoding/decoding. Then the measurement is repeated after MP3 encoding/decoding. The next section discusses the methodology and equipment configuration for our measurements.

V. Methodology

The goal is to measure the watermark contamination in the music signal after blind demixing with ICA. First a baseline is produced that measures the watermark residue in the music channel when the MP3 codec is not employed, in Fig. 10 without the dashed block. Then the residue level is measured by introducing the MP3 codec to the linearly mixed signals as in Fig. 10 with the dashed block. To control the relative contributions of the watermark and music signals, the linear mixing matrix was parameterized for a single mixture angle, **Ø**. In both cases, the latter are demixed using ICA to recover permuted and scaled versions of the originals (Music' and Watermark').

To facilitate simultaneous linearity and residue measurements, the watermark signal and music signals were implemented as a pair of dual tones. The Watermark signal is two equal amplitude tones at 900 Hz and 1000 Hz and two equal amplitude tones for the Music signal are at 2100 Hz and 2250 Hz. The linear dynamic range is determined by measuring the magnitude of the intermodulation (IM) products generated by the tone pair, relative to the magnitude of the respective tone pairs. As an example, for the two tone watermark signal at 900 and 1000 Hz, any 1st order IM products would occur at 800 and 1100 Hz.

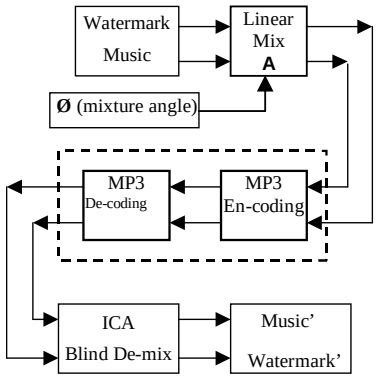

Fig. 10. Mixing and demixing procedure model with/without MP3 codec block (dashed block).

The MP3 codec was implemented using commercial software (Syntrillium, 2000). The following MP3 encoding options were chosen and held constant through all the experiments: 44.1 kHz sample rate, 128 kbps per channel, and stereo (no joint stereo coding).

As a check, the software MP3 decoder was checked with consumer hardware. MP3 files were burned on CDROMs, played back on a Riovolt MP3/CD Player through an RCA SA-155 Integrated Stereo Amplifier (Radio Shack catalog #31-5000) with spectrum measurements obtained from a spectrum analyzer (HP 3588A with opt. 001). The nominal 80 dB of linear dynamic range measured in the MP3 software codec was also observed with the hardware.

VI. Results

An illustration of the results of demixing for a mixture angle of 30 degrees in the absence of the MP3 codec is shown in Fig. 11. It shows more than 80 dB of isolation between the wanted signals and the residue signals. The left column is Watermark' channel with the wanted watermark tones at 900 Hz and 1000 Hz, and the right column is Music' channel with the wanted music tones at 2100 Hz and 2250 Hz.

Fig. 12 is the summary of minimum magnitude differences in dB for the case when the MP3 codec (the dashed block in Fig. 10) was absent and the mixture angle was varied from 0 to 90 degrees at 0.1-degree increments in the top of pictures.

In the bottom of the pictures in Fig. 12 is a summary for the case when the MP3 codec (the dashed block in Fig. 10) was present, but the residues were evaluated only from 0 degree to 90 degree at every 1 degree increments. In this figure, the left column shows the minimal magnitude differences in dB between the residue signal and the wanted signal in the Watermark' channel, and in the same way the right column shows the minimal magnitude differences in Music' channel.

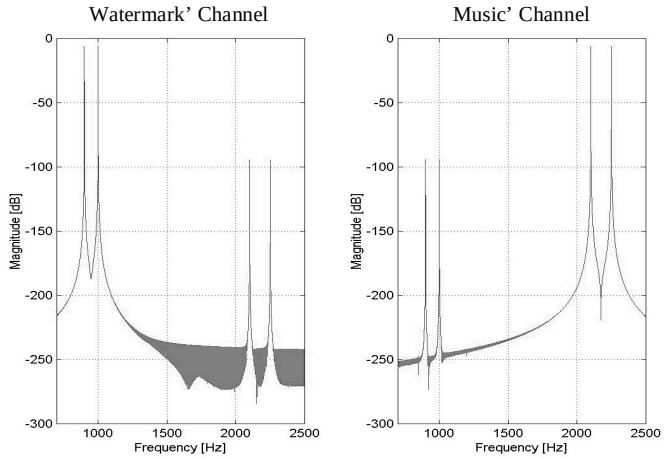

Fig. 11. The residues in both demixed channels.

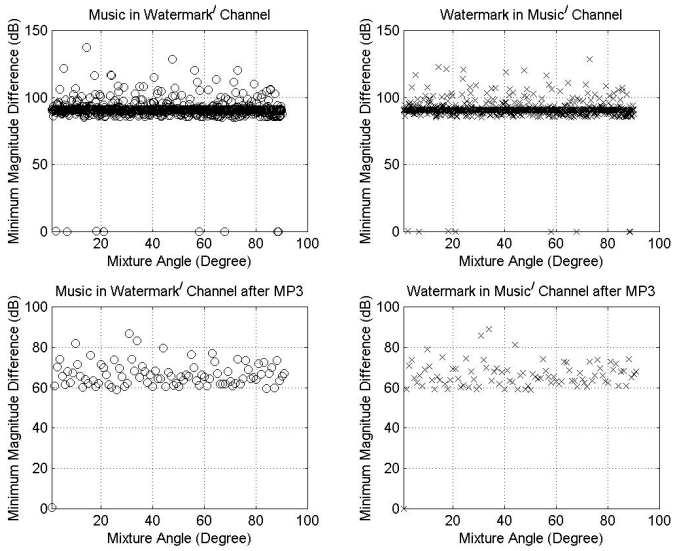

Fig. 12. Minimum magnitude differences (dB) between the residue signal and the wanted signal are shown at each mixture angle from 0 to 90 degrees.

VII. Discussion

Without the MP3 codec, the results of blind demixing with ICA showed watermark attenuation in the music channel was more than 82 dB for a wide range of mixture angles, close to the 90 dB dynamic range of MP3 encoded music. When the linearly mixed signals were passed through the lossy MP3 codec, the residues were approximately 70 dB down from the wanted signals as measured at the 30 and 60 degrees mixture angles. However, when we measured from 0 to 90 degrees at every 1 degree, the residues were approximately larger than 59 dB down. Despite the 23 dB performance deficiency with the MP3 codec, the watermark contamination in the music channel is likely to be imperceptible by the human auditory system. The high quality of the music recovery suggests that there is value in restricted computer codes for the restoration of music from watermarked MP3 digital music files.

The lossy MP3 coding format might affect the statistics of both the watermark and music signals when selected frequency components are nulled. However, in our evaluations with the selected simple tonal pairs representing the watermark and acoustic signals, the adverse affects on blind demixing with ICA are small.

Some open research opportunities include recovery of the absolute amplitude and phase of the music signal. With ICA blind demixing an unknown scale and permutation in the recovered outputs generally occurs. A scale change could also manifest itself as a 180-degree phase shift in the music channel. For stereo music, this may destroy the "depth" perception. If some of the properties of the watermark are known *a priori*, then recovery of the watermark might serve as a guide to recover the music channel's absolute amplitude and sense of phase.

VIII. Summary and Conclusions

This paper examines the application of independent component analysis (ICA), via unsupervised neural networks, to authenticity protection for multimedia products. The blind demixing capability of these neural networks extends signal processing from a one-sensor approach to a multi-sensor approach.

For watermark security, we propose a covert watermarking signal that serves as a vaccination against a dormant digital bacterium. Removal of the watermark triggers the bacterium, which degrades the quality of the product being pirated. We show that our digital-bacteria model meets established technical and ethical requirements for beneficial virus-like programs.

We apply our novel ICA watermarking method to digitally encoded acoustic music files. A watermarked signal was linearly mixed with a music signal. The resulting mixture pair was put through an MP3 codec, and then blindly demixed with ICA to recover the music signal. The watermark contamination in the music channel was measured both when the MP3 codec was absent and when it was present.

Our experimental results show that watermark contamination in the music channel is likely to be inaudible, and that the adverse affects on blind demixing with ICA are small. We conclude that our approach can provide a flexible, robust, and secure system for protecting the authenticity of multimedia products.

References

Amari, S., Chichocki, A., & Yang, H. (1996). A new learning algorithm for blind signal separation. In D. Touretzky, M. Mozer, & M. Hasselmo, *Advances in neural information processing systems 8*. Cambridge, MA: MIT Press.

Barnett, R. (1999). Digital watermarking: applications, techniques, and challenges. *Electronics and Communication Engineering Journal*, 11(4), 173-183.

Bell, A. & Sejnowski, T. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7, 1129-1159.

Bell, A. & Sejnowski, T. (1996). Learning the higher-order structure of a natural sound. *Network: Computation in Neural Systems*, 7, 261-266.

Bontchev, V. (1994). Are “good” computer viruses still a bad idea? *Proceedings of EICAR*, London.

Brandenburg, K. & Popp, H. (2000). An introduction to MPEG Layer-3. *EBU Technical Review*, 1-15.

Cohen, F. (1984). Computer viruses - theory and experiments. *Computers and Security*, 6(1), 22-35.

Comon, P. (1994). Independent component analysis, a new concept?. *Signal Processing*, 36(3), 287-314.

Gonzalez-Serrano, F., Molina-Bulla, H. & Murillo-Fuentes, J. (2001). Independent component analysis applied to digital image watermarking. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (pp. 1997-2000) Salt Lake City.

Hartung, F. & Kutter, M. (1999). Multimedia watermarking techniques. *Proceedings of the IEEE*, 87(7), 1079-1107.

Hyvärinen, A., & Oja, E. (1999). *Independent component analysis: a tutorial*. Tutoial Notes, International Joint Conference on Neural Networks, Washington, DC.

Jessop, P. (1999). The business case for audio watermarking. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (pp. 2077-2074) Phoenix.

Kopriva, I. & Szu, H. (2003). Blind inversion in nonlinear space-variant imaging by using cauchy machine. *Proceedings of the SPIE Conference on Independent Component Analyses, Neural Nets and Wavelets*, Orlando.

Noel, S. & Szu, H. (2000). Multimedia authenticity with ICA watermarks. *Proceedings of Wavelet Applications VII* (pp. 175-184) Orlando.

Oja, E. & Karhunen, J. (1995). Signal separation by nonlinear hebbian learning. *Proceedings of the International Conference on Neural Networks* (pp.417-421) Perth.

Pan, D. (1995). A tutorial on MPEG/audio compression. *IEEE Multimedia Journal*, 60-74.

Quan, A., Szu, A., & Markovitz, Z. (2000). Local ICA for the most wanted face recognition. *Proceedings of Wavelet Applications VI* (pp. 539-551) Orlando.

Seok, J. & Hong, J. (2001). Audio watermarking for copyright protection of digital audio data. *Electronics Letters*, 37(1), 60-61.

Swanson, M., Kobayashi, M., & Tewfick, A. (1998), Multimedia data-embedding and watermarking technologies. *Proceedings of the IEEE*, 86(6), 155-178.

Syntrillium Software Corporation (2003). Syntrillium software - home of Cool Edit 2000 and Cool Edit Pro. Web page, http://www.syntrillium.com/. Last accessed March 21, 2003.

Szu, H. (1999). ICA-enabling techniques for intelligent sensory processing. *IEEE Circuits and Systems Newsletters*. 14-41.

Szu, H. (1999). Progress in unsupervised artificial neural networks for image demixing applications. *IEEE Industrial Electronics Society Newsletter*, 46(2). 7-12.

Yim, S.-B., Willey, J., Landa, J., & Szu, H. (2003). Watermarking MP3 encoded music and ICA blind demixing. *Proceedings of the SPIE Conference on Independent Component Analyses, Neural Nets and Wavelets*, Orlando.

Yu, D. & Satter, F. (2002). Robust image watermarking based on independent component analysis. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (pp. 4177-4177) Orlando.

Yu, D. Satter, F., & Ma, K.-K. (2002). Watermark detection and extraction using ICA method. *EURASIP Journal on Applied Signal Processing*, 2002(1), 92-104.